

Fig. S4. Human performances and PROBE model fit with 4 recurrent action sets. Shaded lines, performances from 30 healthy subjects (16 females, aged 18-30 y/o) in recurrent episodes plotted against the number of trials following episodes onsets. Shaded areas are s.e.m. across subjects (detailed legend in **Fig. 1**). The experimental session consisted of 24 recurrent episodes identical to that from Exp. 1 (see text), except that four mappings between stimuli and correct responses re-occured pseudo-randomly across episodes. The four mappings were fully incongruent. Note that subjects performed as in open episodes in Exp. 1 (see **Fig. 1**) with no peaks of mutual dependence of successive decisions in the first trials of episodes. **Lines** +/- **error bars** (mean +/- s.e.m.), performances predicted by the fitted PROBE model (details in **Fig. 2**): correct and exploratory response rates were computed in every trial according to the actual history of subjects's responses. Mutual dependence of successive correct decisions predicted by the model was computed as the mutual information between two successive correct decisions produced by the model independently of actual subjects' responses (1 simulation for each subject). Best-fitting model parameters (mean(s.e.m.)): inverse temperature β=35(2.3); noise ε=0.04(.003); bound N=3.4(.3); learning rate α=0.34(.04); recollection entropy η=0.75(.03); confirmation bias θ=0.34(.06). Note that the parameters are close to those from Exp. 1 (see **Table S1**) See **Supporting Information** (section Comments on model fits) for additional comments regarding model and subjects' behavior.